








COVID-19 case prediction via wastewater surveillance in a low-prevalence urban community: a modeling approach

Yifan Zhu ^{a,†}, Wakana Oishi^{b,†}, Chikako Maruo ^b, Sewwandi Bandara ^a, Mu Lin^a, Mayuko Saito ^c, Masaaki Kitajima ^d and Daisuke Sano ^d ^{a,b,e,*}

^a Department of Frontier Sciences for Advanced Environment, Graduate School of Environmental Studies, Tohoku University, Aoba 6-6-06, Aramaki, Aoba-ku, Sendai, Miyagi 980-8597, Japan

^b Department of Civil and Environmental Engineering, Graduate School of Engineering, Tohoku University, Aoba 6-6-06, Aramaki, Aoba-ku, Sendai, Miyagi 980-8597, Japan

^c Department of Virology, Graduate School of Medicine, Tohoku University, 2-1 Seiryō-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan

^d Division of Environmental Engineering, Faculty of Engineering, Hokkaido University, North 13 West 8, Kita-ku, Sapporo, Hokkaido 060-8628, Japan

^e Research Institute for Humanity and Nature, 457-4 Kamigamo Motoyama, Kita Ward, Kyoto 603-8047, Japan

*Corresponding author. E-mail: daisuke.sano.e1@tohoku.ac.jp

[†]These authors contributed equally to this work.

 YZ, 0000-0002-6754-3590; CM, 0000-0001-8812-1840; SB, 0000-0001-5669-8602; MS, 0000-0003-4622-7027; DS, 0000-0001-8075-6972

ABSTRACT

Estimating and predicting the epidemic size from wastewater surveillance results remains challenging for the practical implementation of wastewater-based epidemiology (WBE). In this study, by employing a highly sensitive detection method, we documented the time series of SARS-CoV-2 RNA occurrence in the wastewater influent from an urban community with a 360,000 population in Japan, from August 2020 to February 2021. The detection frequency of the viral RNA increased during the outbreak events of COVID-19 and the highest viral RNA concentration was recorded at the beginning of January 2021, amid the most serious outbreak event during the study period. We found that: (1) direct back-calculation still suffers from great uncertainty dominated by inconsistent detection and the varying gap between the observed wastewater viral load and the estimated patient viral load, and (2) the detection frequency correlated well with reported cases and the prediction of the latter can be carried out via data-driven modeling methods. Our results indicate that wastewater virus occurrence can contribute to epidemic surveillance in ways more than back-calculation, which may spawn future wastewater surveillance implementations.

Key words: COVID-19 surveillance, data-driven modeling, epidemic prediction, wastewater-based epidemiology

HIGHLIGHTS

- Time series of SARS-CoV-2 wastewater presence in a low-prevalence community.
- SARS-CoV-2 detection frequency correlates with reported cases.
- Reported cases can be predicted by detection frequency within a period.
- Data-driven methods may facilitate wastewater-based epidemic modeling.

1. INTRODUCTION

The COVID-19 pandemic has caused over 5 million deaths as of 16 November 2021 (WHO 2021) and is still deeply interrupting the global society. In a new era when communities are adapting to revive, there is a substantial need for a surveillance tool that can identify both asymptomatic and symptomatic patients with high efficiency, preferably also having a short turnaround time, wide coverage, non-invasive, anonymous in nature, and with a low implementation cost.

Previous studies have confirmed the persistent shedding of SARS-CoV-2 viral RNA in the feces and urine from patients during the infection course (Wölfel *et al.* 2020) and the viral shedding at the asymptomatic stage (Tang *et al.* 2020; Zhang *et al.* 2020). A review on viral shedding demonstrated that the shedding lasts 14–28 days (Jones *et al.* 2020). Detections of SARS-CoV-2 viral RNA from the wastewater have been reported in a number of studies, including in low-prevalence areas (Haramoto *et al.* 2020; Randazzo *et al.* 2020; D'Aoust *et al.* 2021; Hata *et al.* 2021) and prior to the surge of an outbreak (Peccia *et al.* 2020). These studies provided a clear rationale to monitor the viral components in wastewater to perceive

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

the spread of disease in a given catchment. Improved detection sensitivity had been reported by using the solid fraction of wastewater (Peccia *et al.* 2020; Balboa *et al.* 2021; Kitamura *et al.* 2021), and the continuous detection of viral RNA in the suspended solid rather than the supernatant throughout the surveillance period was reported (Kitamura *et al.* 2021). Pre-amplification of the target sequence is one option to improve sensitivity in RT-PCR assay. A previous study reported a 100-fold improvement of quantification limit in an RT-PCR assay for coronavirus in clinical tests (Lau *et al.* 2003). Therefore, it may also be suitable for wastewater samples that may contain a low amount of target genetic materials.

Although much has already been achieved regarding concentrating and quantifying SARS-CoV-2 from wastewater, wastewater surveillance implementations remain limited, mainly due to a lack of means to interpret the result. The efforts of converting the results into prevalence level, often referred to as ‘back-calculation’, initiated with the global emergence of detection reports. Being able to quantitatively interpret the wastewater surveillance outcome can greatly facilitate the situation assessment and countermeasure formation (Sims & Kasprzyk-Hordern 2020; Thompson *et al.* 2020). The mass balance function commonly seen in applications regarding drug and pesticide monitoring had been adopted by some recent studies. Briefly, the measured biomarker concentration is multiplied by the wastewater flow rate to obtain the total load (Ahmed *et al.* 2020a; Hart & Halden 2020; Fernandez-Cassi *et al.* 2021; Hasan *et al.* 2021; Saththasivam *et al.* 2021), which can then be divided by an average load to obtain an approximate shedder base. Besides, coefficients such as the degradation factor and the shedding ratio are often included to improve the estimation. However, as research went deeper, it was realized that the shedding profile of infected individuals is rather erratic and cannot be accurately represented by a fixed rate (Walsh *et al.* 2020; Buonerba *et al.* 2021), and other sources of uncertainty may further bring down the credibility of prevalence estimated in this manner (Li *et al.* 2021).

Unquantifiable data points did not receive sufficient consideration either. The wastewater virus concentration sometimes exceeds the limit of detection (LoD) or sensitivity threshold, but not the limit of quantification (LoQ), especially in the early stage of an epidemic or in a low-prevalence region. Data points that fall into this range can only be identified as ‘positive’ without a quantifiable value, making them binomial (either positive or negative). This creates a grey area of data as the previous mass balance approach fails to apply to data that fall within this range. Thus, it is critical to develop alternative modeling approaches that manage to extract information from these data, yet related studies are scarce.

This study aimed to analyze the relationship between the number of reported COVID-19 cases and the RT-qPCR-based viral RNA detection in wastewater influent. To achieve this, we monitored the time series of SARS-CoV-2 RNA concentrations in wastewater influent from Sendai city, Japan, from August 2020 to February 2021, with a highly sensitive detection method that consists of the recovery of viral RNA from wastewater solids and pre-amplification of cDNA before qPCR assay. Then, using the detection data, we present two metrics that help researchers extract information from the wastewater surveillance results. Specifically, we applied and evaluated two modeling approaches that apply to quantifiable and binomial wastewater virus detections, respectively.

2. METHODS

2.1. Sample collection

A total of 51 influent wastewater samples were collected from a municipal wastewater plant (WWTP) in Sendai, Miyagi, Japan that receives approximately 69% of wastewater generated in the city. Samples were taken twice a week, at 10 a.m., on Tuesday and Thursday from August 2020 to February 2021, from an influent line that serves the major urban area and about 360,000 people. All samples were grab samples (250 mL). Samples were immediately transported to the laboratory after collection and stored at -80°C until analysis.

2.2. Virus concentration, nucleic acid extraction, and cDNA synthesis

SARS-CoV-2 RNA was recovered from 40 mL influent wastewater sample. The suspended solid was concentrated by centrifugation at 5,000 g for 10 min at 4°C . After the supernatant was removed, 1 mL of TRIzol reagent (Thermo Fisher Scientific, MA, USA) was added to the concentrated suspended solid, then the suspension was homogenized using a vortex mixer. The total volume of the suspension was less than 3.5 mL. A 140 μL aliquot of the concentrate was processed for viral RNA extraction using the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) following the manufacturer’s instructions, and the viral RNA was eluted in 60 μL of elution buffer provided in the kit. The whole process recovery of the SARS-CoV-2 RNA was verified based on the concentration of Pepper Mild Mottle Virus (PMMoV) in wastewater samples, and 140 μL aliquot was extracted from samples both before and after concentration. A 10 μL sample of extracted RNA was used to obtain 20 μL of

cDNA with the high-capacity cDNA RT Kit (Thermo Fisher Scientific). To synthesize cDNA for SARS-CoV-2, the random primer included in the kit was substituted with CDC nCOV_N1-R Primer (CN 10006831, Integrated DNA Technologies, Inc., Iowa, USA) (10 μ M) (CDC 2021).

2.3. SARS-CoV-2 genome pre-amplification and quantification by qPCR assay

The PCR-based pre-amplification method was applied to cDNA prior to the qPCR assay. The pre-amplification was performed with *TaKaRa Ex Taq*[®] Hot Start Version (Takara Bio, Kusatsu, Japan) and the CDC nCOV_N1 Primers (CN10006830 and CN10006831, Integrated DNA Technologies, Inc.) (CDC 2021). Each 50 μ L reaction contained 20 μ L of cDNA, 0.25 μ L of *TaKaRa Ex Taq HS* (5 U/ μ L) (Takara Bio), 5 μ L of 10 \times Ex Taq Buffer (Mg²⁺ plus) (20 mM) (Takara Bio), 4 μ L of dNTP Mixture (Takara Bio), and 400 nM of forward and reverse primers. The PCR cycling condition was 2 min at 94 $^{\circ}$ C, followed by 10 cycles of 30 s at 94 $^{\circ}$ C, 30 s at 55 $^{\circ}$ C, and 1 min at 72 $^{\circ}$ C. The pre-amplification step was simultaneously applied to 18 or 20 μ L of standard DNA (2.0 to 2.0₁₀₄ copies/ μ L) created via 10-fold dilution series of 2019-nCoV_N_PositiveControl (CN10006625, Integrated DNA Technologies, Inc.). The pre-amplification step was also applied to 20 μ L of TE buffer as the negative control for the pre-amplification and qPCR.

The concentrations of SARS-CoV-2 and PMMoV viral RNA were determined by real-time qPCR on a CFX96 real-time PCR detection system (Bio-Rad, Hercules, CA, USA). The amplification reaction was performed with SsoAdvanced Probes Supermix (Bio-Rad). For SARS-CoV-2 viral RNA, we used the same forward/reverse primers from the previous pre-amplification step with nCOV_N1 Probe Aliquot (CN10006832, Integrated DNA Technologies, Inc.) (CDC 2021), and RT-qPCR was performed only on pre-amplified samples. Each 20 μ L reaction contained 5 μ L of pre-amplified cDNA, 10 μ L of SsoAdvanced Probes Supermix (Bio-Rad), 500 nM of forward and reverse primers, and 200 nM of fluorogenic probe. The PCR cycling condition was 30 s at 95 $^{\circ}$ C, followed by 40 cycles of 10 s at 95 $^{\circ}$ C and 30 s at 60 $^{\circ}$ C. The number of SARS-CoV-2 genome copies was determined by a standard curve generated with the pre-amped standard samples (1.0 \times 10¹ to 1.0 \times 10⁵ copies/reaction). Each sample was quantified in triplicate. The amplification efficiency in the real-time PCR was at least 80%.

The concentration of SARS-CoV-2 viral RNA in the influent wastewater sample, $C_{v,w}$ (copies/mL), was determined by the following equation:

$$C_{v,w} = 1,000 \times \frac{V}{V_w} \times \frac{V_{f,ex}}{V_{s,ex}} \times \frac{V_{f,syn}}{V_{s,syn}} \times \frac{V_{f,pre}}{V_{s,pre}} \times C_{qPCR} \quad (1)$$

where V is the volume of concentrated wastewater suspended with TRIzol reagent (mL) (1.1–3.5 mL), C_{qPCR} is the concentration of cDNA applied to qPCR (copies/ μ L), V_w is the volume of raw wastewater (mL). $V_{f,ex}$, $V_{s,ex}$, $V_{f,syn}$, $V_{s,syn}$, $V_{f,pre}$, and $V_{s,pre}$ are all the volumes of samples in the intermediate steps (μ L). Subscripts f and s stand for the final and starting volumes, while subscripts ex , syn , and pre stand for the RNA extraction, the cDNA synthesis, and the pre-amplification, respectively. The LoQ was 118–375 copies/mL based on the quantification limit in a qPCR assay (2 copies/ μ L of pre-amped standard samples) and Equation (1).

2.4. PMMoV genome quantification by qPCR assay

For the quantification of PMMoV viral RNA, the cDNA was applied to real-time qPCR without pre-amplification. The amplification reaction was performed with SsoAdvanced Probes Supermix (Bio-Rad), the reverse and forward primers, and probe. Each 20 μ L reaction contained 5 μ L of cDNA, 10 μ L of SsoAdvanced Probes Supermix (Bio-Rad), 500 nM of forward and reverse primers, and 200 nM of fluorogenic probe. The PCR cycling condition used for the detection of CDC N1 was also used for PMMoV. The number of PMMoV genome copies of one reaction was determined by a standard curve generated with the standard samples (1.0 \times 10²–1.0 \times 10⁶ copies/reaction). Each sample was quantified in triplicate.

The concentration of PMMoV viral RNA in influent wastewater samples was determined by the following equation:

$$\text{PMMoV, copies/mL} = 1,000 \times \frac{V_{f,ex}}{V_{s,ex}} \times \frac{V_{f,syn}}{V_{s,syn}} \times C_{qPCR} \quad (2)$$

where C_{qPCR} is the concentration of cDNA applied to qPCR (copies/ μ L).

2.5. Viral load calculation

For data points with quantified SARS-CoV-2 RNA concentration, we applied a model developed in a prior study (Equation (3)) (Zhu *et al.* 2021). Briefly, patients are assigned different shedding rates according to a pre-defined shedding function that takes into account how many days the patients are into the infection course. The transition of the shedding rate is governed by a matrix whose dimension is determined by the patient report coverage and the shedding duration of infected individuals. Thus, with a consecutive patient number record, the viral load excreted by the active shedder base on a given day can be calculated as follows:

$$L_{p[d]} = \sum_{n=1}^j P_{[n,d]} * C_{f,n} * r_f * m_f \quad (3)$$

In this model, $L_{p[d]}$ is the patient viral load on calendar day d , j is the shedding duration, $P_{[n,d]}$ represents the number of patients on the n th day of the infection course on day d and can be called from a matrix used to store and track the infection status of reported cases. $C_{f,n}$ is the fecal shedding rate (copies/g feces) that is calculated from a separate function, r_f is the ratio of patients who develop fecal shedding, and m_f stands for the daily feces load (g/day). More details can be found in Zhu *et al.* (2021).

The number of daily COVID-19 cases in Sendai city was obtained from the Sendai municipal government. The case report covers the entire population of Sendai city (about 1 million) while the catchment area only covers 360,000 people, the patient viral load was thus multiplied by 0.3.

The patient viral load L_p was then compared to the wastewater viral load L_w calculated by Equation (4).

$$L_{w[d]} = C_{v,w[d]} * Q_d * k_1 \quad (4)$$

In this equation, the measured concentration of SARS-CoV-2 RNA in the wastewater $C_{v,w}$ is multiplied by the wastewater flow rate Q and a dilution coefficient k_1 to obtain the wastewater viral load L_w . The daily flow rate was acquired from the treatment plant operator. A dilution coefficient k_1 was introduced to help normalize the wastewater viral load, the reason being that toilet flushing is not evenly distributed throughout the day. Previous studies have reported that a peak of toilet flushing in the morning accounts for about 25% of the total daily flushes (Butler *et al.* 1995; Campisano & Modica 2015). However, due to a lack of in-depth information, k_1 was set to 1 in this study.

2.6. Positive rate modeling

To test whether wastewater surveillance can provide information about how the epidemic may unfold, for data points with positive yet unquantifiable results, a prediction model framework was established (Figure 1). First, the correlation between the reported cases and the positive rate was assessed by Spearman's rank-order correlation and generalized linear model (GLM). Rolling two- and four-week were used to ensure enough data points in a calculation window. The positive rate was calculated as the number of positive signals divided by the total sample number in the given calculation window (rolling two- or four-week). Then, the positive rates from consecutive calculation windows were used as inputs to predict the reported cases in the last calculation window. Both the positive rate and reported cases were assigned to the last week of the calculation window.

Three models, GLM, artificial neural network (ANN), and random forest (RF), were employed to perform the prediction tasks for their ability to solve nonlinear regression problems and learn from available data. For each model, a pre-determined portion of the dataset (80%) was randomly selected for model training; the trained models were then used to perform prediction on the remaining part of the dataset (testing data). This random sampling-training prediction process was repeated 5,000 times. The mean squared error (MSE) of actual value versus predicted value was calculated each time for performance evaluation. The optimal number of inputs was determined through a performance analysis (Supplementary Table S2). The data pre-treatment, model configuration, prediction, and statistical analysis were all performed using the R programming language; the related code is provided in the Supplementary material.

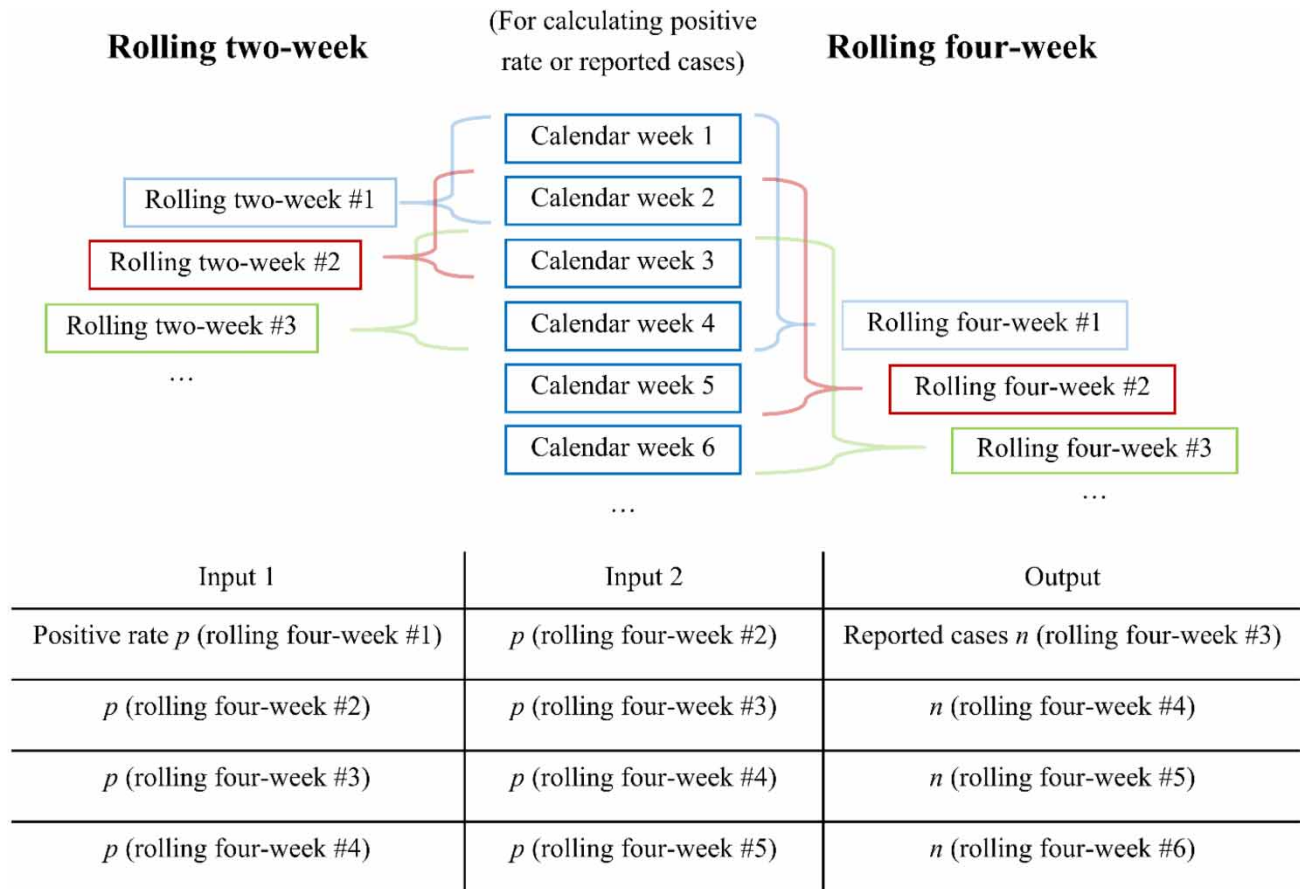


Figure 1 | A brief illustration of the model framework. The positive rate and cumulative cases in a rolling week are calculated from a calculation window of two or four weeks, the values are assigned to the last week included in the calculation window. For example, when using rolling two-week, the positive rate during calendar weeks 1 and 2 is denoted as p (rolling week #1). In the prediction models, the model inputs are the positive rates in consecutive calculation windows (in chronological order), while the output is the cumulative cases of the following calculation window.

3. RESULTS

3.1. COVID-19 prevalence summary

As a low-prevalence region, Sendai was not severely hit by COVID-19 during the study period. A total of 2,142 cases were reported in Sendai city from August 3, 2020, to February 28, 2021, and can be approximately assigned into two outbreak events (Figure 2). The first one lasted between late October and late November 2020. The peak appeared on October 27, 2020, when 38 patients were reported. The second outbreak event that struck between mid-December and late January was more critical with a higher daily case count. There were 8 days when the daily reported cases exceeded the peak in the first outbreak, and 63 cases were reported on January 14, 2021, marking an all-time high. It is worth mentioning that although the daily reported patient number had a temporal dip during the New Year holiday, it was more likely due to the reduced testing capacity and delayed reporting rather than actual easing of epidemic. Following the second outbreak event, the daily reported cases dropped to a low level and remained that way until the end of the study period.

3.2. Wastewater surveillance summary

The concentration of PMMoV RNA in wastewater influent ranged from $5.2 \log_{10}$ to $5.8 \log_{10}$ genome copies/mL, while in concentrated influent samples ranged from $5.4 \log_{10}$ to $6.4 \log_{10}$ genome copies/mL. As the occurrence of PMMoV RNA in concentrated wastewater samples was stable, this may suggest there was no significant loss of SARS-CoV-2 RNA in the whole quantification process.

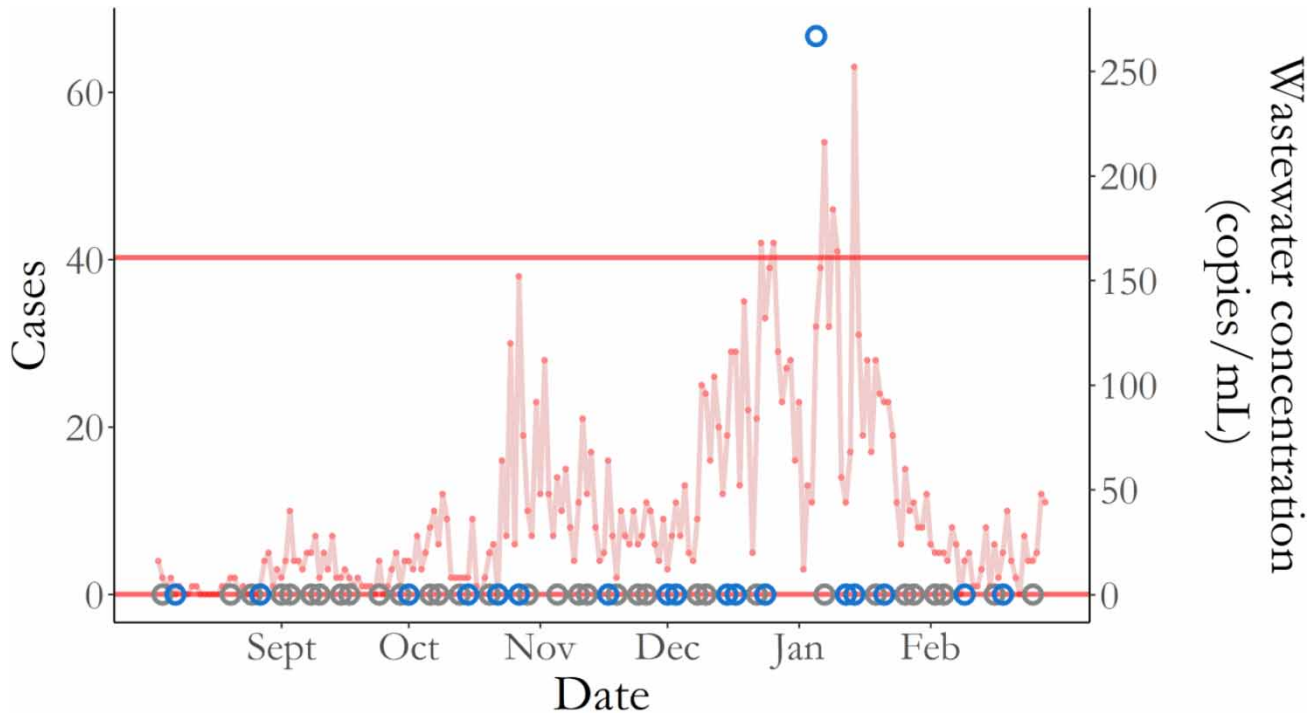


Figure 2 | The time series of the SARS-CoV-2 RNA occurrence in wastewater influent and daily reported cases (red line and points). Positives are shown as blue circles, while negatives are grey. The upper and lower red horizontal lines represent the LoQ (median) and qualitative sensitivity, respectively.

In total, 51 samples were examined throughout the surveillance period, and 33 of them were negative. The genome concentration corresponding to the highest C_t value in our assay (referred to as qualitative sensitivity hereafter) was estimated to be 0.025 copies/mL by extrapolating the standard curve. Seventeen samples recorded C_t values greater than this, but still lower than the LoQ, ranging from 1.18×10^2 to 3.75×10^2 copies/mL with a median of 1.61×10^2 copies/mL. The amplification efficiency in real-time PCR was from 80 to 120% which was the acceptable range according to the MIQE guidelines (Bustin *et al.* 2009). The coefficient of determination of the standard curves was greater than 0.99 in each assay. No PCR products were detected in negative controls. We considered samples that were tested positive in at least one well in triplicate analysis as positive.

During the study period, the measured viral RNA concentration exceeded the LoQ only once on January 5, 2021 (Figure 2). With a concentration of 2.67×10^2 copies/mL, the daily wastewater viral load estimated from Equation (4) would be 6.74×10^{13} copies. However, the viral load contributed by reported patients on that day, calculated using Equation (3), would be 3.40×10^9 copies with 203 cumulative cases in the 26-day patient viral load calculation window, meaning a L_w/L_p value of 1.98×10^4 . Also, this quantifiable signal occurred prior to the peak of reported cases which came 9 days later, on January 14, 2021. The wastewater virus concentration did not exceed the LoQ again despite a higher daily case count reported in the following days.

On the other hand, due to the lack of quantifiable data points, non-quantitative detection gave us a more consistent dataset to work on. The first positive signal occurred on August 7, 2020, with just 21 cumulative cases in the patient viral load calculation window and an estimated patient viral load of 3.42×10^8 copies, which translates into a theoretical wastewater virus concentration of only 2.70×10^{-3} copies/mL, far below the qualitative sensitivity. However, assuming the concentration sits somewhere between the qualitative sensitivity and the LoQ, the wastewater viral load would have a range of 6.32×10^9 – 4.06×10^{13} , thus a L_w/L_p range of 1.85×10^1 – 1.19×10^5 , covering the L_w/L_p value estimated from the quantifiable detection on January 5, 2021. Over the study period, a total of 18 (35.29%) samples tested positive. Although a positive signal does not directly translate into wastewater viral concentration, consecutive positives may indicate a high viral load with higher confidence. In that sense, two consecutive positives appeared four times and all of them occurred during the two outbreak events.

3.3. Positive rate modeling

A stronger correlation was found between the four-week positive rate and cumulative cases than that between the two-week positive rate and cumulative cases (Figure 3). Therefore, the prediction models were used to predict the cumulative cases using the four-week positive rate. By testing different amounts of input numbers, we found the optimal value was two (Supplementary Table S2). Among the three models (Figure 4), ANN offered the best overall performance (median MSE: 7520.22) followed by the other two (median MSE: 9,038.60 (GLM), 12,021.26 (RF)). For about half of the data points (45.83%, 11 in 24), the actual four-week cumulative cases were within the 95% CI range of the prediction. For the remaining data points, the average error was 17.51%.

4. DISCUSSION

By employing a highly sensitive detection method, we monitored the time series of SARS-CoV-2 RNA occurrence in wastewater influent from an urban community with a population of 360,000. Eighteen out of the 51 influent samples yielded positive signals, and 17 samples had SARS-CoV-2 RNA concentrations lower than the LoQ. By examining the reported cases, we found the positive rate of detection has a strong correlation (four-week rolling window, $\rho=0.7598$, $p<0.05$) with the cumulative cases in the same time frame and established prediction models, hoping to extend the knowledge on wastewater-based epidemiology (WBE) implementation strategies.

In this study, the LoQ was around 1.61×10^2 copies/mL, about 1 log above that reported in the previous studies (Ahmed *et al.* 2020a; Hokajärvi *et al.* 2021). This is due to the decreased volume of samples throughout the analysis in this study. Specifically, the volume of a sample input was approximately 1/10, 1/6, and 1/10 of the total volume of a suspension obtained in a previous step, in RNA extraction, cDNA synthesis, and qPCR, respectively. The LoQ may be improved by using other different extraction kits that use a larger volume of samples for extraction. Nevertheless, the method we employed is a feasible option because the data is obtained within 1 day, and our analysis has not been restricted by the stock shortages of manufacturers.

We used PMMoV as an internal process control for SARS-CoV-2 detection from suspended solids. PMMoV is detected throughout the year and is abundant in wastewater (10^6 – 10^{10} copies/L) (Kitajima *et al.* 2014; Symonds *et al.* 2018; Ahmed

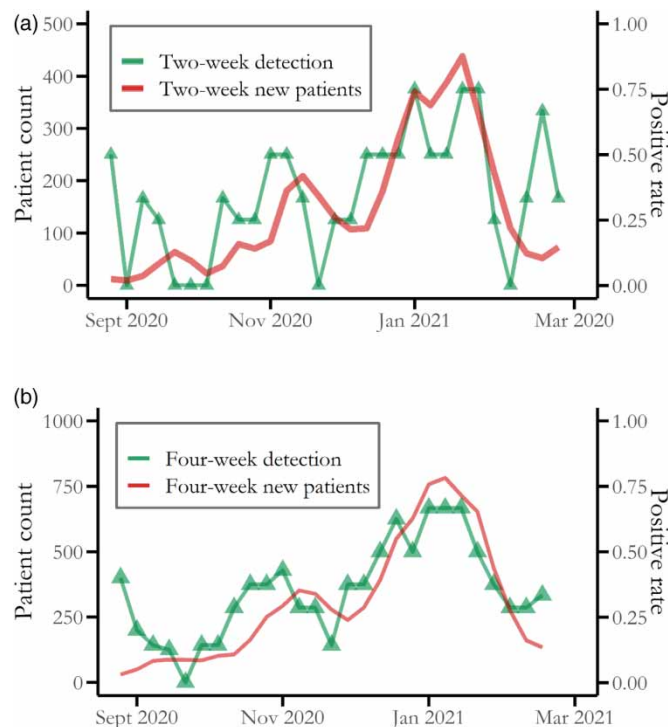


Figure 3 | The time series of positive rate and cumulative cases. (a) The calculation window is rolling two-week. (b) The calculation window is rolling four-week. Spearman's rank-order correlation coefficients: 0.4996 (two-week, $p<0.05$) and 0.7598 (four-week, $p<0.05$).

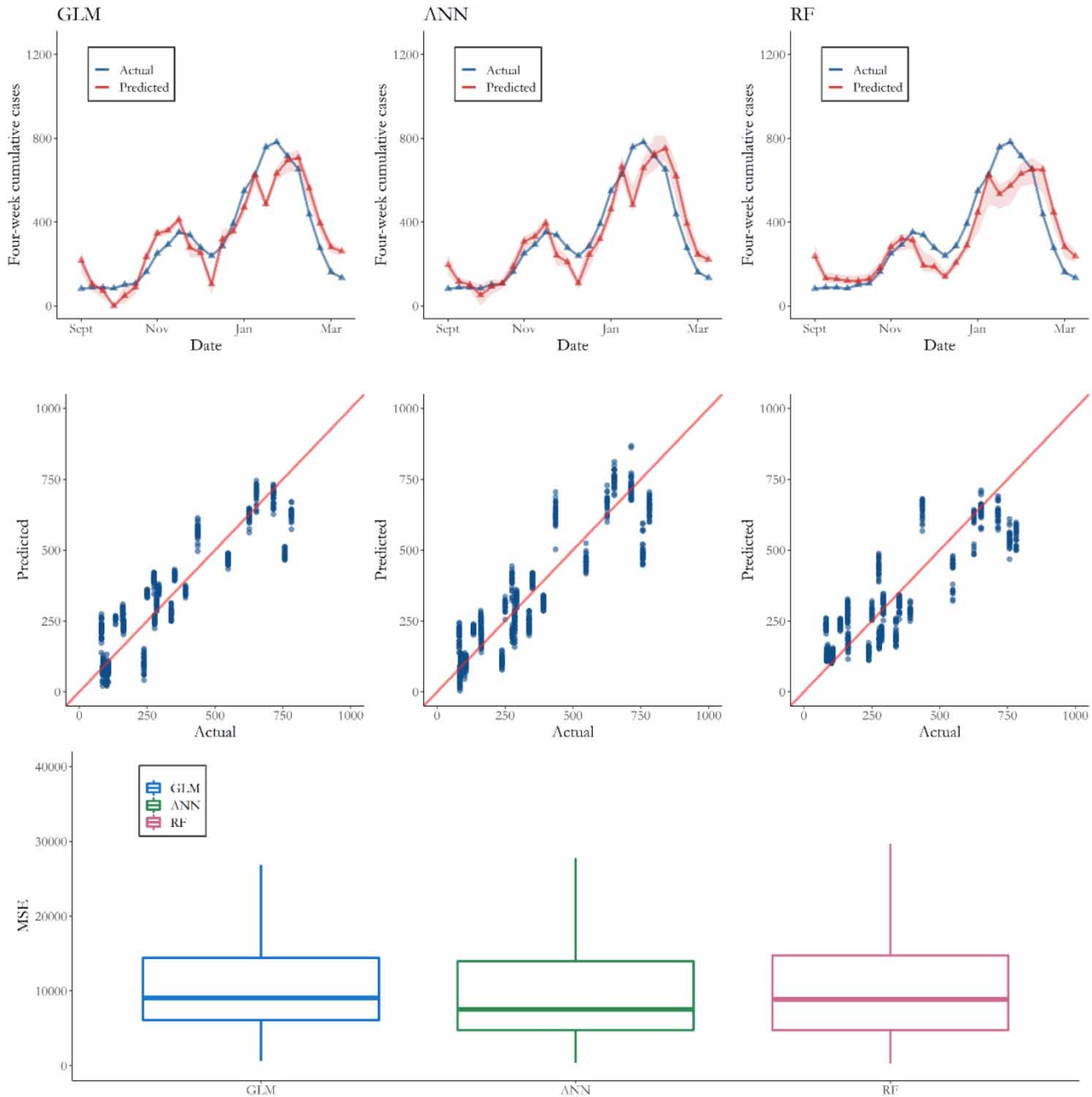


Figure 4 | The cumulative cases predicted by the three models. The inputs were the positive rates in two consecutive rolling four-weeks, while the output was the cumulative cases within the following four-week. Normalization was applied to inputs and output before model training and all data were denormalized back to the original scale once the prediction was conducted. The three rows of figures are the time series of actual versus predicted (median and 95% CI) four-week cumulative cases for each model, the scatter plot of 1,000 randomly selected pairs of actual and predicted values, and the boxplot of MSE distribution from the 5,000 predictions of each model, respectively.

et al. 2020b), which may allow for using it as an internal control of RT-qPCR for a wastewater sample (Haramoto *et al.* 2020). A previous study reported that the recovered load of PMMoV correlated with that of murine hepatitis virus, suggesting that PMMoV is the potential indicator of the efficiency of SARS-CoV-2 (Torii *et al.* 2022). We concluded that there was no significant loss throughout the analysis because the PMMoV concentration was consistent in both raw and concentrated influent samples. Future studies should decide the best whole process control for extraction from suspended solids. The better options are human coronaviruses 229E and HKU1 although their longitudinal concentration has not been reported (Bibby *et al.* 2011; Bibby & Peccia 2013).

The pre-amplification employed in this study increases the number of amplicons in the downstream qPCR. The theoretical qualitative sensitivity should have the same Ct value as that obtained from one copy per reaction, which in this study is 32.9. However, this value was surpassed by multiple samples whose Ct values reached up to 40.0. A possible explanation for the results was that organic compounds in the influent samples inhibited the amplification efficiency in PCR. We evaluated the effect of inhibitors using a commercially available RNA positive control according to the Gibson *et al.* (2012) study but did not observe lowered efficiency in pre-amplification and qPCR of the positive control RNA, indicating that other reasons were responsible for the greater-than-expected Ct values. It may be explained by the different affinity efficiency of primers and polymerase to the target amplicons between the cDNA derived from viral RNA and plasmid DNA used as the positive control.

From the perspective of early warning, getting a positive signal from wastewater can be a solid proof that the virus has started circulating in the community (Randazzo *et al.* 2020; Fernandez-Cassi *et al.* 2021). So far, different studies have reported varied sensitivity. Hong *et al.* (2021) reported that a positive signal in hospital wastewater requires 253–409 positive cases out of 10,000 individuals, while Hata *et al.* (2021) detected the presence of SARS-CoV-2 RNA in municipal wastewater when the number of cases was <1.0 per 100,000 people, and Betancourt *et al.* (2021) reported a positive detection when there were only one symptomatic and two asymptomatic individuals among a total of 311 residents in a student dormitory. In practice, the varied detection sensitivity can be mainly attributed to the different experimental methods used as well as the characteristics of the sewage system from which samples are collected. A standardized method may contribute to the comparison and integration of studies (Weidhaas *et al.* 2021). In this study, when the first positive signal was recorded, the number of active shedders estimated from the clinical reports was only 21 in the catchment area with about 360,000 people. Nevertheless, a bigger active shedder base does not guarantee a positive signal in subsequent detections. Even during the summit of the second outbreak event which enabled the only signal above LoQ, negatives were still recorded. Such inconsistent detection had also been reported in other studies (Fernandez-Cassi *et al.* 2021; Hong *et al.* 2021), adding another layer of complexity.

Despite the strong interest in quantitative wastewater surveillance, a streamlined solution has yet to be formed. Especially, although the experimental side has received substantial attention which led to more sensitive and reliable detection, the analytical side still lacks adequate investigation and verification. There is a noteworthy knowledge gap in how to associate the measured wastewater virus concentration with the epidemic size in the catchment area. So far, most studies had tried directly correlating the abundance of viral RNA with reported cases. However, the following are the findings from our study: (1) a positive signal occurred when the speculated active shedder group was supposedly far from large enough to enable a successful detection, (2) higher reported cases did not translate to high wastewater virus concentration, and (3) there is a significant yet uncertain gap between the observed wastewater viral load and the viral load contributed by the supposed active shedder base, presented as L_w/L_p in this study, all point to a conclusion that at the current stage, the uncertainty associated with the wastewater viral load is still a great hindrance to reliable back-calculation.

The dimensionless metric L_w/L_p largely determines the robustness of back-calculation. Although a stable L_w/L_p is ideal for back-calculation and was therefore assumed in some recent studies, its value seems to be both time- and location-specific due to various factors. For instance, in Sendai, August and September are the rainy season, because the major urban area is served by a combined sewer, which likely aggravated the dilution of viral RNA and led to a lower L_w/L_p . This is supported by the lower bound of L_w/L_p estimated for August 7, 2020 when the first positive signal appeared. The way patient viral load L_p was calculated also implies it can be impacted by societal factors. For instance, a high level of underreporting may occur under limited testing capacity, leading to a smaller speculated active shedder base, thus a smaller L_p and a larger L_w/L_p . Similarly, if the asymptomatic infection ratio increase, a larger L_w/L_p can also be expected. Knowing this, some critical epidemic-related information may be drawn by keeping a close eye on L_w/L_p . Nevertheless, it should be pointed out that our calculations of L_w and L_p were based on a set of assumptions including the shedding profile, which may be further refined once more medical evidence becomes available.

As shown in this study, in wastewater surveillance projects, researchers may obtain positive yet unquantifiable signals, especially in a low-prevalence period/region. As far as we know, no wastewater surveillance study has utilized binominal data other than as occurrence indicator yet. But, the detection frequency, or positive rate, might serve as a suitable indicator of the virus occurrence upon which further analysis can be performed. This indicates that binominal results may also be utilized to help with epidemic surveillance while establishing a precise connection between the wastewater viral load and the prevalence level remains challenging and entails further research. Rolling four-week was used as the calculation window in this study, but it may be shortened by increasing the sampling frequency or the number of samples collected each time, albeit

more time- and resource-consuming. It should be noted, though, that using a positive rate as an indicator may only be feasible in a low-prevalence region or at the early stage of an outbreak event. There exists an upper bound of epidemic size beyond which its linear correlation with positive rate becomes invalid, which may explain why the epidemic peaks were not successfully modeled in this study. On the other hand, a higher prevalence level means a higher chance of getting quantifiable signals, and back-calculation models should take over once developed.

When a causal relationship between input and output is difficult to establish, data-driven methods like those used in this study may be employed. However, being data-driven also means training data need to be accumulated to fine-tune the model, and prediction does not always match the reality. With all the uncertainties, it should be reiterated that wastewater surveillance ought not to be a stand-alone tool and its outcome should be interpreted along with other information sources before reaching any conclusion. For instance, in the early stage of an epidemic when clinical testing capacity is often compromised, wastewater detection may be put into action quicker and cover a larger area. Nevertheless, our study shows that a positive rate may be an important indicator, as also recognized by a recent study (Fernandez-Cassi *et al.* 2021). Also, in terms of prediction accuracy, as stated above, environmental and societal factors may affect the detection result; thus, adding explanatory variables into the model may improve the model performance.

Several limitations of this study should be noted. First, the lag between symptom onset and reporting was not included in modeling. Counting in the delay in case reporting may explain the 9-day delay between the quantified virus concentration and the case peak, it may also improve the correlation between the positive rate and cumulative cases, as cases would be assigned to an earlier date. However, existing studies on the delay between symptom onset and hospitalization had varied estimations ranging from 7 days (Huang *et al.* 2020) to a much shorter 1.2 days (Lauer *et al.* 2020). Therefore, without enough information about the local testing and reporting practice, integrating this factor into the model may introduce further error. Second, grab samples are prone to short-term heterogeneity of viral RNA abundance, which may affect the representativeness of samples. While composite samples collected by an autosampler may improve the consistency of detection, the viral RNA may also get highly diluted as toilet flushing mainly occurs during certain times, resulting in false negatives. Designing a sampling strategy that captures the toilet flushing peak, therefore, may be a viable solution as suggested by recent studies (Betancourt *et al.* 2021).

5. CONCLUSIONS

As the world is still under the shadow of COVID-19, on top of timely medical and societal intervention, each and every tool that helps monitor the situation and alerts the society is worth looking into. In this study, using a highly sensitive assay, we (1) monitored the occurrence of SARS-CoV-2 viral RNA in the wastewater of an urban area in Japan for over 7 months and (2) established a model framework to help extend the existing knowledge base about analyzing and interpreting the surveillance results. Particularly, we found that although quantitative epidemic size estimation based on measured virus concentration is still challenging, the positive rate of wastewater virus detection is strongly correlated with reported cases and can be used for its prediction, which may guide toward novel wastewater surveillance strategies. Our findings may not only strengthen the application of wastewater surveillance in the current COVID-19 pandemic but also help the scientific community prepare for other public health challenges.

ACKNOWLEDGEMENTS

This research was supported by the Japan Agency for Medical Research and Development (AMED) under Grant No. JP20wm0125001 and by JST SPRING under Grant Number JPMJSP2114. We appreciate the generous cooperation of the Sendai City Construction Bureau for providing wastewater samples.

REFERENCES

- Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V. & Mueller, J. F. 2020a [First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community](#). *Science of the Total Environment* **728**, 138764. doi:10.1016/j.scitotenv.2020.138764.
- Ahmed, W., Kitajima, M., Tandukar, S. & Haramoto, E. 2020b [Recycled water safety: current status of traditional and emerging viral indicators](#). *Current Opinion in Environmental Science and Health* **16**, 62–72. doi:10.1016/j.coesh.2020.02.009.

- Balboa, S., Mauricio-Iglesias, M., Rodriguez, S., Martínez-Lamas, L., Vasallo, F. J., Regueiro, B. & Lema, J. M. 2021 *The fate of SARS-COV-2 in WWTPS points out the sludge line as a suitable spot for detection of COVID-19*. *Science of the Total Environment* **772**, 145268. doi:10.1016/j.scitotenv.2021.145268.
- Betancourt, W. Q., Schmitz, B. W., Innes, G. K., Prasek, S. M., Pogreba Brown, K. M., Stark, E. R., Foster, A. R., Sprissler, R. S., Harris, D. T., Sherchan, S. P., Gerba, C. P. & Pepper, I. L. 2021 *COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention*. *Science of the Total Environment* **779**, 146408. doi:10.1016/j.scitotenv.2021.146408.
- Bibby, K. & Peccia, J. 2013 *Identification of viral pathogen diversity in sewage sludge by metagenome analysis*. *Environmental Science and Technology* **47** (4), 1945–1951. doi:10.1021/es305181x.
- Bibby, K., Viau, E. & Peccia, J. 2011 *Viral metagenome analysis to guide human pathogen monitoring in environmental samples*. *Letters in Applied Microbiology* **52** (4), 386–392. doi:10.1111/j.1472-765X.2011.03014.x.
- Buonerba, A., Corpuz, M. V. A., Ballesteros, F., Choo, K. H., Hasan, S. W., Korshin, G. V., Belgiorno, V., Barceló, D. & Naddeo, V. 2021 *Coronavirus in water media: analysis, fate, disinfection and epidemiological applications*. *Journal of Hazardous Materials* **415**. doi:10.1016/j.jhazmat.2021.125580.
- Bustin, S. A., Benes, V., Garson, J. A., Hellems, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W. & Shipley, G. L. 2009 *The MIQE Guidelines: Minimum information for publication of quantitative real-time PCR experiments*. *Clinical Chemistry* **55**, 611–622.
- Butler, D., Friedler, E. & Gatt, K. 1995 *Characterising the quantity and quality of domestic wastewater inflows*. *Water Science and Technology* **31** (7), 13–24. doi:10.1016/0273-1223(95)00318-H.
- Campisano, A. & Modica, C. 2015 *Appropriate resolution timescale to evaluate water saving and retention potential of rainwater harvesting for toilet flushing in single houses*. *Journal of Hydroinformatics* **17** (3), 331–346. doi:10.2166/hydro.2015.022.
- CDC 2021 *Real-Time RT-PCR Primers and Probes for COVID-19*. CDC. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html> (accessed 29 November 2021).
- D'Aoust, P. M., Mercier, E., Montpetit, D., Jia, J. J., Alexandrov, I., Neault, N., Baig, A. T., Mayne, J., Zhang, X., Alain, T., Langlois, M. A., Servos, M. R., MacKenzie, M., Figeys, D., MacKenzie, A. E., Graber, T. E. & Delatolla, R. 2021 *Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in communities with low COVID-19 incidence and prevalence*. *Water Research* **188**, 116560. doi:10.1016/j.watres.2020.116560.
- Fernandez-Cassi, X., Scheidegger, A., Bänziger, C., Cariti, F., Tuñas Corzon, A., Ganesanandamoorthy, P., Lemaitre, J. C., Ort, C., Julian, T. R. & Kohn, T. 2021 *Wastewater monitoring outperforms case numbers as a tool to track COVID-19 incidence dynamics when test positivity rates are high*. *Water Research* **200**, 117252. doi:10.1016/j.watres.2021.117252.
- Gibson, K. E., Schwab, K. J., Spencer, S. K. & Borchardt, M. A. 2012 *Measuring and mitigating inhibition during quantitative real time PCR analysis of viral nucleic acid extracts from large-volume environmental water samples*. *Water Research* **46** (13), 4281–4291. doi:10.1016/j.watres.2012.04.030.
- Haramoto, E., Malla, B., Thakali, O. & Kitajima, M. 2020 *First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan*. *Science of the Total Environment* **737**, 140405. doi:10.1016/j.scitotenv.2020.140405.
- Hart, O. E. & Halden, R. U. 2020 *Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: feasibility, economy, opportunities and challenges*. *Science of the Total Environment* **730**, 138875. doi:10.1016/j.scitotenv.2020.138875.
- Hasan, S. W., Ibrahim, Y., Daou, M., Kannout, H., Jan, N., Lopes, A., Alsafar, H. & Yousef, A. F. 2021 *Detection and quantification of SARS-CoV-2 RNA in wastewater and treated effluents: surveillance of COVID-19 epidemic in the United Arab Emirates*. *Science of the Total Environment* **764**, 142929. doi:10.1016/j.scitotenv.2020.142929.
- Hata, A., Hara-Yamamura, H., Meuchi, Y., Imai, S. & Honda, R. 2021 *Detection of SARS-CoV-2 in wastewater in Japan during a COVID-19 outbreak*. *Science of the Total Environment* **758**, 143578. doi:10.1016/j.scitotenv.2020.143578.
- Hokajärvi, A. M., Rytönen, A., Tiwari, A., Kauppinen, A., Oikarinen, S., Lehto, K. M., Kankaanpää, A., Gunnar, T., Al-Hello, H., Blomqvist, S., Miettinen, I. T., Savolainen-Kopra, C. & Pitkänen, T. 2021 *The detection and stability of the SARS-CoV-2 RNA biomarkers in wastewater influent in Helsinki, Finland*. *Science of the Total Environment* **770**, 145274. doi:10.1016/j.scitotenv.2021.145274.
- Hong, P. Y., Rachmadi, A. T., Mantilla-Calderon, D., Alkahtani, M., Bashawri, Y. M., Al Qarni, H., O'Reilly, K. M. & Zhou, J. 2021 *Estimating the minimum number of SARS-CoV-2 infected cases needed to detect viral RNA in wastewater: to what extent of the outbreak can surveillance of wastewater tell us?* *Environmental Research* **195**, 110748. doi:10.1016/j.envres.2021.110748.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J. & Cao, B. 2020 *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. *The Lancet* **395** (10223), 497–506. doi:10.1016/S0140-6736(20)30183-5.
- Jones, D. L., Baluja, M. Q., Graham, D. W., Corbishley, A., McDonald, J. E., Malham, S. K., Hillary, L. S., Connor, T. R., Gaze, W. H., Moura, I. B., Wilcox, M. H. & Farkas, K. 2020 *Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19*. *Science of the Total Environment* **749**, 141364. DOI: 10.1016/j.scitotenv.2020.141364
- Kitajima, M., Iker, B. C., Pepper, I. L. & Gerba, C. P. 2014 *Relative abundance and treatment reduction of viruses during wastewater treatment processes – identification of potential viral indicators*. *Science of the Total Environment* **488–489** (1), 290–296. doi:10.1016/j.scitotenv.2014.04.087.
- Kitamura, K., Sadamasu, K., Muramatsu, M. & Yoshida, H. 2021 *Efficient detection of SARS-CoV-2 RNA in the solid fraction of wastewater*. *Science of the Total Environment* **763**, 144587. doi:10.1016/j.scitotenv.2020.144587.

- Lau, L. T., Fung, Y. W. W., Wong, F. P. F., Lin, S. S. W., Wang, C. R., Li, H. L., Dillon, N., Collins, R. A., Tam, J. S. L., Chan, P. K. S., Wang, C. G. & Yu, A. C. H. 2003 A real-time PCR for SARS-coronavirus incorporating target gene pre-amplification. *Biochemical and Biophysical Research Communications* **312** (4), 1290–1296. doi:10.1016/j.bbrc.2003.11.064.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G. & Lessler, J. 2020 The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine* **172** (9), 577–582. doi:10.7326/M20-0504.
- Li, X., Zhang, S., Shi, J., Luby, S. P. & Jiang, G. 2021 Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. *Chemical Engineering Journal* **415**, 129039. doi:10.1016/j.cej.2021.129039.
- Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., Weinberger, D. M., Arnold, W. & Omer, S. B. 2020 Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology* **38**, 1164–1167. doi:10.1038/s41587-020-0684-z.
- Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A. & Sánchez, G. 2020 SARS-CoV-2 RNA in wastewater anticipated COVID-19 increase in a low prevalence area. *Water Research* **181**. doi:10.1016/j.watres.2020.115942.
- Saththasivam, J., El-Malah, S. S., Gomez, T. A., Jabbar, K. A., Remanan, R., Krishnankutty, A. K., Ogunbiyi, O., Rasool, K., Ashhab, S., Rashkeev, S., Bensaad, M., Ahmed, A. A., Mohamoud, Y. A., Malek, J. A., Abu Raddad, L. J., Jeremijenko, A., Abu Halaweh, H. A., Lawler, J. & Mahmoud, K. A. 2021 COVID-19 (SARS-CoV-2) outbreak monitoring using wastewater-based epidemiology in Qatar. *Science of the Total Environment* **774**, 145608. doi:10.1016/j.scitotenv.2021.145608.
- Sims, N. & Kasprzyk-Hordern, B. 2020 Future perspectives of wastewater-based epidemiology: monitoring infectious disease spread and resistance to the community level. *Environment International* **139**, 105689. doi:10.1016/j.envint.2020.105689.
- Symonds, E. M., Nguyen, K. H., Harwood, V. J. & Breitbart, M. 2018 Pepper mild mottle virus: a plant pathogen with a greater purpose in (waste)water treatment development and public health management. *Water Research* **144**, 1–12. doi:10.1016/j.watres.2018.06.066.
- Tang, A., Tong, Z., Wang, H., Dai, Y., Li, K., Liu, J., Wu, W., Yuan, C., Yu, M., Li, P. & Yan, J. 2020 Detection of novel coronavirus by RT-PCR in stool specimen from asymptomatic child, China. *Emerging Infectious Diseases* **26** (6), 1337–1339. doi:10.3201/eid2606.200301.
- Thompson, J. R., Nancharaiyah, Y. V., Gu, X., Lee, W. L., Rajal, V. B., Haines, M. B., Girones, R., Ng, L. C., Alm, E. J. & Wuertz, S. 2020 Making waves: wastewater surveillance of SARS-CoV-2 for population-based health management. *Water Research* **184**. doi:10.1016/j.watres.2020.116181.
- Torii, S., Oishi, W., Zhu, Y., Thakali, O., Malla, B., Yu, Z., Zhao, B., Arakawa, C., Kitajima, M., Hata, A., Ihara, M., Kyuwa, S., Sano, D., Haramoto, E. & Katayama, H. 2022 Comparison of five polyethylene glycol precipitation procedures for the RT-qPCR based recovery of murine hepatitis virus, bacteriophage phi6, and pepper mild mottle virus as a surrogate for SARS-CoV-2 from wastewater. *Science of the Total Environment* **807**, 150722. doi:10.1016/j.scitotenv.2021.150722.
- Walsh, K. A., Jordan, K., Clyne, B., Rohde, D., Drummond, L., Byrne, P., Ahern, S., Carty, P. G., O'Brien, K. K., O'Murchu, E., O'Neill, M., Smith, S. M., Ryan, M. & Harrington, P. 2020 SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *Journal of Infection* **81** (3), 357–371. doi:10.1016/j.jinf.2020.06.067.
- Weidhaas, J., Aanderud, Z. T., Roper, D. K., VanDerslice, J., Gaddis, E. B., Ostermiller, J., Hoffman, K., Jamal, R., Heck, P., Zhang, Y., Torgersen, K., Laan, J., Vander & LaCross, N. 2021 Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Science of the Total Environment* **775**, 145790. doi:10.1016/j.scitotenv.2021.145790.
- WHO 2021 WHO Coronavirus (COVID-19) Dashboard. WHO. Available from: <https://covid19.who.int/> (accessed 16 November 2021).
- Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., Niemeyer, D., Jones, T. C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brünink, S., Schneider, J., Ehmann, R., Zwirgmaier, K., Drosten, C. & Wendtner, C. 2020 Virological assessment of hospitalized patients with COVID-2019. *Nature* **581** (7809), 465–469. doi:10.1038/s41586-020-2196-x.
- Zhang, J. C., Wang, S. B. & Xue, Y. D. 2020 Fecal specimen diagnosis 2019 novel coronavirus-infected pneumonia. *Journal of Medical Virology* **92** (6), 680–682. doi:10.1002/jmv.25742.
- Zhu, Y., Oishi, W., Saito, M., Kitajima, M. & Sano, D. 2021 Early warning of COVID-19 in Tokyo via wastewater-based epidemiology: how feasible it really is? *Journal of Water and Environment Technology* **19** (3), 170–183. doi:10.2965/jwet.21-024.

First received 28 July 2021; accepted in revised form 30 January 2022. Available online 7 February 2022